ISRN TELCORDIA--2012-07+PR-0GARAU
**Prepared for: Office of Naval Research**

# Fast Multiscale Algorithms for Information Representation and Fusion

## Technical Progress Report No. 7

Devasis Bassu, Principal Investigator

Contract: N00014-10-C-0176

Applied Communication Sciences

One Telcordia Drive

Piscataway, NJ 08854-4157

April 2012

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**APR 2012** | 2. REPORT TYPE | 3. DATES COVERED |
| 4. TITLE AND SUBTITLE<br>**Fast Multiscale Algorithms For Information Representation And Fusion** | | 5a. CONTRACT NUMBER |
| | | 5b. GRANT NUMBER |
| | | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | | 5d. PROJECT NUMBER |
| | | 5e. TASK NUMBER |
| | | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Applied Communication Sciences,One Telcordia Drive,Piscataway,NJ,08854** | | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | 10. SPONSOR/MONITOR’S ACRONYM(S) |
| | | 11. SPONSOR/MONITOR’S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited.**

13. SUPPLEMENTARY NOTES
**The original document contains color images.**

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | | **14** | |

# 1 Abstract

In the seventh quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) continued design/implementation of the Multiscale Singular Value Decomposition (SVD) algorithm and c) packaging for releasing the software as open source. This report documents experimental results with the Multiscale SVD algorithms.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and wrap up development of the multiscale SVD algorithms. No problems are currently anticipated.

# Table of Contents

# 2 Summary

In this quarter, we continued design and implementation of the new multiscale SVD (MSVD) algorithms. We applied the MSVD to a publicly available LIDAR dataset for the purposes of distinguishing between vegetation and the forest floor. The initial findings are presented in this report.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and wrap up development of the multiscale SVD algorithms. No problems are currently anticipated.

# 3 Introduction

The primary project effort over the last quarter focused on completing the design and continued development of the multiscale SVD algorithms [1]. Preliminary results from experiments conducted on a publicly available LIDAR dataset [5] are provided in Section 5.

# 4 Methods, Assumptions and Procedures

## 4.1 Multiscale Singular Value Decomposition

The Multiscale Singular Value Decomposition (MSVD) was introduced in the earlier technical report [6]. The MSVD provides a spectral readout of the dataset at all scales. Two broad variants of the MSVD algorithm are addressed in this project.

The first MSVD is computed by imposing a dyadic grid on a selected low number of dimensions (say 1, 2 or 3 typically representing temporal, spatial and spatio-temporal dimensions) of some multi-valued dataset. The algorithm is described in section 4.1.1 of the report [6]. An example using this algorithm is described in section 5.1 below.

The second MSVD variant computes the Singular Value Decomposition (SVD) for points contained in balls of various sizes around each point in the data set. The approximate *k*-Nearest Neighbors algorithms developed earlier in this project provides the desired scalability to rapidly select the data points for each scale. For small sized datasets, the exact neighbors may be easily computed. This version of the MSVD algorithm is described in section 4.1.2 of the report [6]. We are currently applying this algorithm to a real-world LIDAR data set. This experiment is described in section 5.2 below.

## 4.2 Deliverables / Milestones

| Date | Deliverables / Milestones | Status |
|---|---|---|
| Oct 2010 | Progress report for period 1, 1st quarter | ✓ |
| Jan 2011 | Progress report for period 1, 2nd quarter / complete randomized matrix decompositions task | ✓ |
| Apr 2011 | Progress report for period 1, 3rd quarter / complete approximate nearest neighbors task | ✓ |
| Jul 2011 | Progress report for period 1, 4th quarter / complete experiments – part 1 | ✓ |
| Oct 2011 | Progress report for period 2, 1st quarter | ✓ |
| Jan 2012 | Progress report for period 2, 2nd quarter / complete multiscale SVD task | ✓ |
| Apr 2012 | Progress report for period 2, 3rd quarter | ✓ |
| Jul 2012 | Progress report for period 2, 4th quarter / complete experiments – part 2 | |
| Oct 2012 | Progress report for period 3, 1st quarter | |
| Jan 2013 | Progress report for period 3, 2nd quarter / complete multiscale Heat Kernel task | |
| Apr 2013 | Progress report for period 3, 3rd quarter | |
| Jul 2013 | Final project report + software + documentation on CDROM / complete experiments – part 3 | |

# 5 Results and Discussion

Two examples are presented below to illustrate

## 5.1 Example 1: Sine Curve (MSVD using a 2-D dyadic grid)

The dataset comprises 629 2-dimensional points ($x$, **sin**($x$)) where $x$ is sampled uniformly from [0, 2π] with 0.01 interval size (see Figure 1).

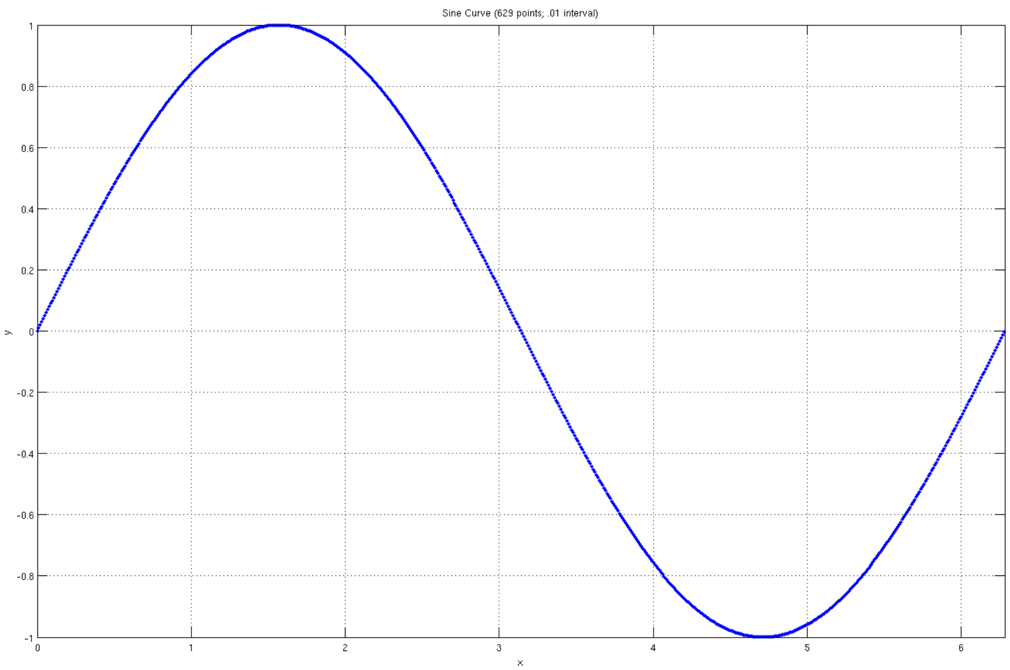


**Figure 1** Example 1: Sine curve dataset

The MSVD is computed by imposing a dyadic grid on the dataset for scales 0 through 4. At scale 0, we have a single rectangular grid – the whole dataset. At scale 1, we have 4 rectangles defined by {[0,π), [π,2π]}×{[-1,0), [0,1]}. The rectangles are sub-divided recursively to obtain grids for higher scales. At each scale, an SVD is computed for all the points inside each rectangle (if the rectangle is not-empty and has enough points). Next, we show the computed singular values and vectors at scales 0 through 4 for the dataset.

Figure 2 shows the singular values for scales 0 through 4 (top-to-bottom). Each row chart depicts the singular values at a given scale. The number of blocks in each row is the number of rectangles at that scale (e.g., scale 0 has only 1, scale 1 has 4, and scale 2 has 16). To map the blocks to Figure 1, first impose the grid, start with the bottom-left rectangle and move up vertically. Once finished, move to the adjacent rectangle on the bottom-right and repeat.

A white/blank block indicates that the rectangle did not have enough points to compute a SVD. Otherwise, each block is a bar chart with 2 sections/colors (corresponding to the proportion of information contributed by each of the two dimensions). It provides visual cues as to whether the data is effectively 1 or 2 dimensions in any given interval (rectangle) and scale. You should now be easily able to locate the two bends of the sine curve at each scale (look for blocks with 2 colors!).

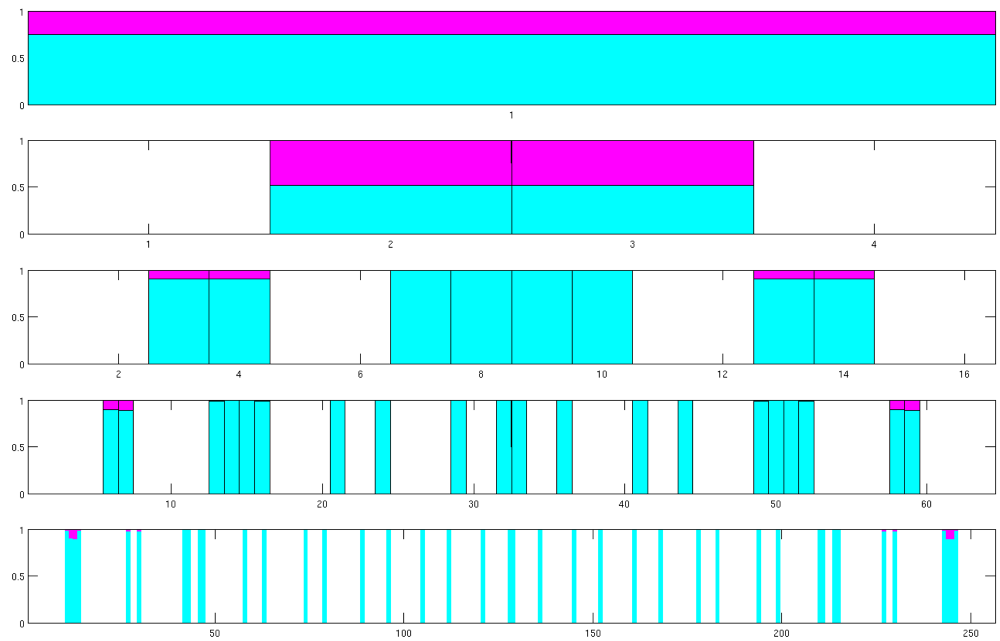

**Figure 2** Example 1: Singular values at scales {0,1,2,3,4} - sine curve dataset

Next, we show the singular vectors associated with the various intervals/scales. The **red** line represents the major (first singular vector) axis while the **green** line represents the minor (second singular vector) axis. The sub-graphs in each plot are placed corresponding to the layout of the 2-dimensional dataset for easy visualization.

Figure 3 and Figure 4 show the computed singular vectors at scales 0 and 1 respectively. As expected, the representation at scale 0 is pretty bad (the dataset is highly non-linear taken as a whole whereas the SVD is suitable for linear structures). While the representation is still not that great for scale 1, it reveals the fact that the data set is localized to the upper-left and lower-right quadrants.

To address the non-linearity of the dataset, we must drill down further to a suitable higher scale where the data is approximately locally linear. Figure 5 shows the singular values at scale 4. At this scale, one can visually see the sine curve. Further, one can trace the curve by simply following the first singular vector (red line).

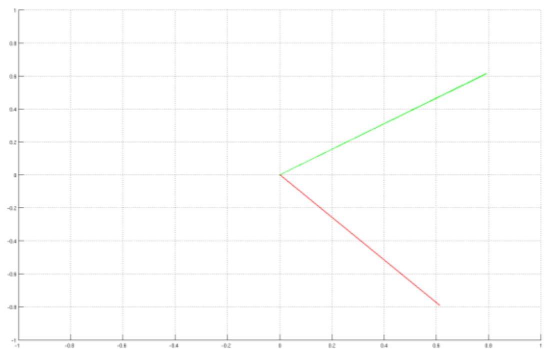

**Figure 3** Example 1: Singular values at scale 0 – sine curve dataset

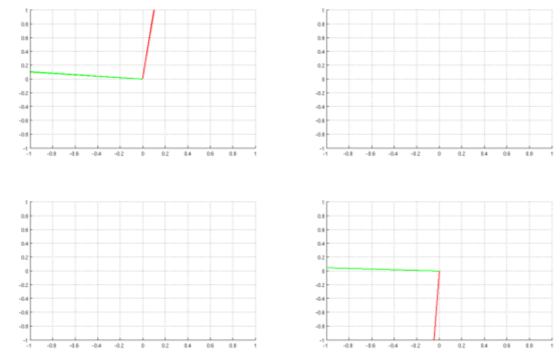

**Figure 4** Example 1: Singular values at scale 1 – sine curve dataset

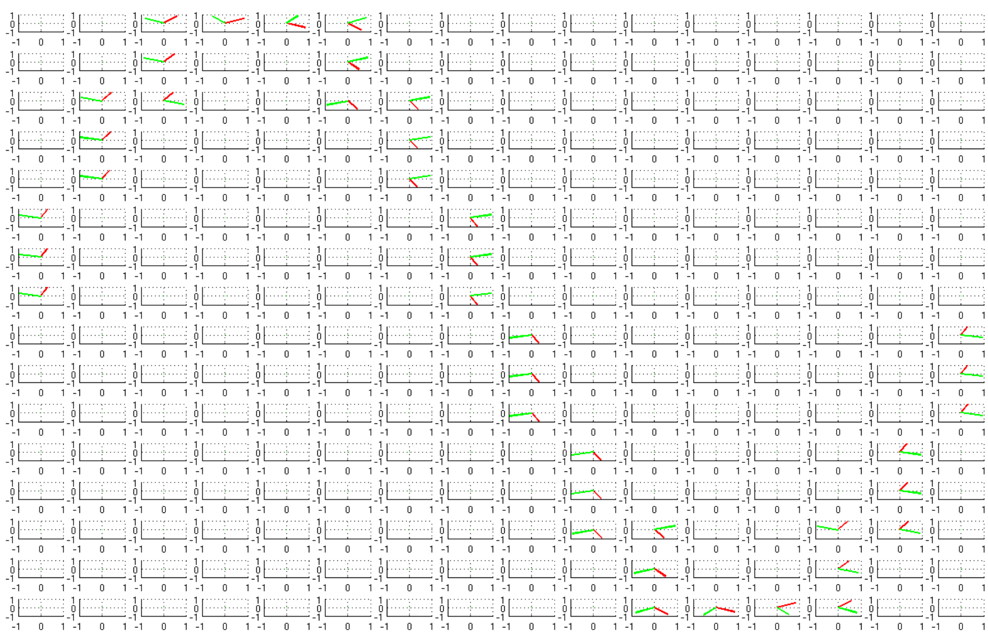

**Figure 5** Example 1: Singular values at scale 4 – sine curve dataset

The dataset is thus characterized by the singular values (singular value proportions are a more effective alternative) and singular vectors for a select set of scales. Each point in the dataset may be simply mapped to the characterization for its parent interval at each scale. For data analysis purposes, one first computes the MSVD for the dataset and then uses this geometric characterization for further analysis (classification, detection, etc.).

## 5.2 Example 2: LIDAR Dataset (MSVD using nearest neighbors)

This publicly available dataset [5] contains LIDAR data representing sections of forest floor and vegetation. An analysis of the dataset for classification purposes is presented in [4]. The dataset comprises 639,520 data points, each categorized as floor or vegetation. Each point is a 3-dimensional spatial position $(x,y,z)$. The dataset is depicted in Figure 6.

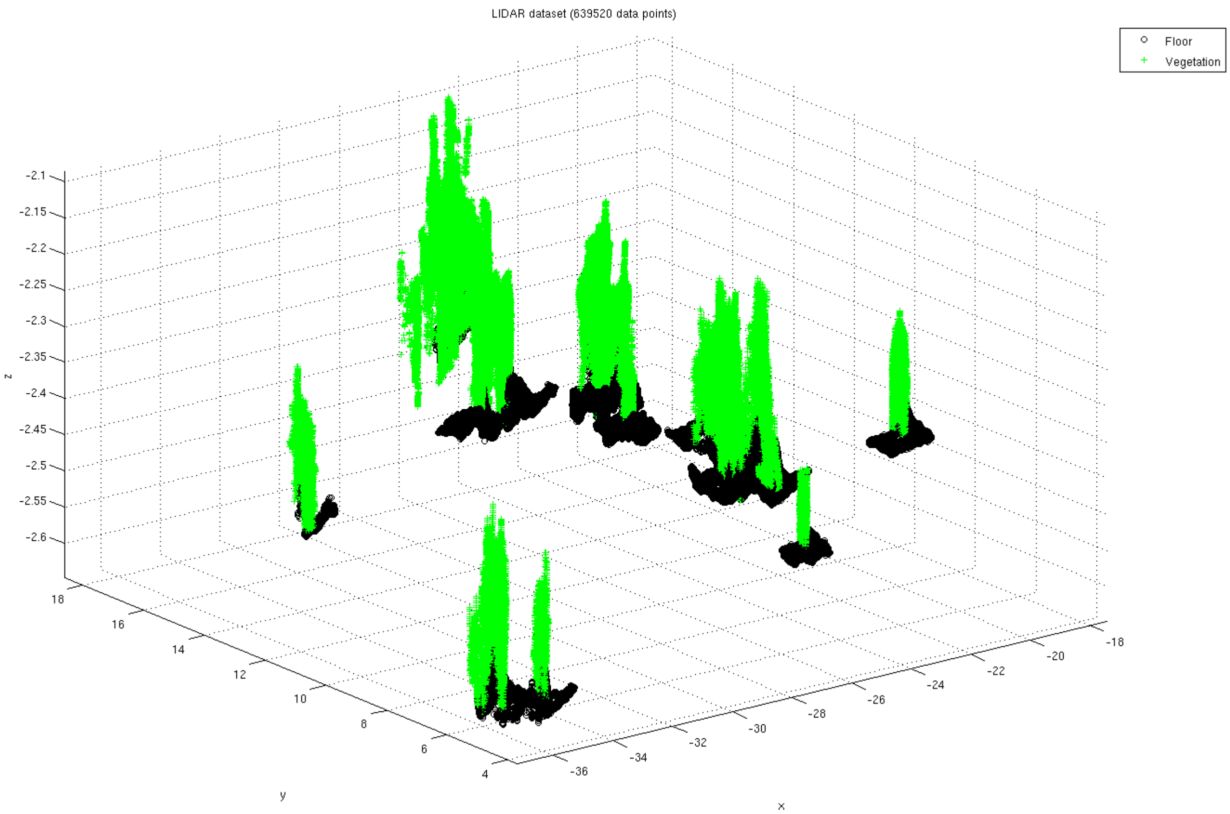


**Figure 6** Example 2: LIDAR dataset

The sensitivity and specificity measures are used to provide metrics for the classification task. The classification accuracy reported in the paper [4] is 95% as *min*{sensitivity, specificity}.

In the first round of experimentation, we used the grid variant of the MSVD algorithm to obtain a characterization for each data point. Various feature sets were selected and the specificity and sensitivity metrics were computed for each selection. The actual classification task was performed using support-vector machines (SVM). The results are listed below.

| Scales | Sensitivity | Specificity |
|---|---|---|
| 2,3,4,5,6 | 77% | 93% |
| 6 | 75% | 96% |
| 5,6 | 84% | 96% |

Note: In every case, the feature set comprises the coordinates ($x$, $y$, $z$) and the singular value proportions and singular vectors.

As expected, the results indicate higher scales (finer grids) capture local geometry better which is central to determining the difference between floor and vegetation. Further, using a combination of scales provides better insight into the local geometry. The relatively low sensitivity value may be explained by noting that the MSVD characterization in the grid variant is truly a geometric characterization of each interval (rectangle) at each scale; *not* necessarily representing each point in the interval. As an example, there may be an interval (at some scale) that contain both floor and vegetation data points. However, using the grid MSVD, all points in that interval have the same characterization attributed to the interval. This would deteriorate the performance of any classifier.

To address this issue, we use the second MSVD variant to compute the local geometry around each point. While this experiment is still ongoing, we report an initial result using 20 approximate nearest neighbors (ANN) disregarding the size of the ball around each point. This is not a significant issue as there are enough points in the vicinity. Using these 20-ANN points, we computed the MSVD for the dataset. For the same test/training sets used in the paper [4], we obtained *min*{sensitivity,specificity}=91%. A 10-discretized version of the SVM pushed the value up to 93%. We will also be computing the MSVD using the exact NN to measure the loss due to the ANN algorithm.

# 6 Conclusions

The project is on track with completed design of the multiscale SVD algorithms. The implemented algorithms are being tried out on a real-world LIDAR dataset with promising preliminary results. We will continue with algorithmic improvements and experimentation using the developed algorithms in the next quarter.

No problems are currently anticipated.

# 7 References

[1] G. Lerman, *Quanitfying curvelike structures of measures by using Jones quantities*, C.P.A.M., vol. 56, issue 8, pages 1294-1365.

[2] G. H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, vol. **2** (2), pages 205–224, 1965.

[3] G.H. Golub, C.F. Van Loan, *Matrix Computations (3rd edition.)*, Johns Hopkins University Press, Baltimore, 1996.

[4] N. Brodu, D. Lague, *3D Terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology*, version 3, 2012. URL: http://arxiv.org/abs/1107.0550

[5] LIDAR dataset URL: http://nicolas.brodu.numerimoire.net/common/recherche/canupo/benchmark.tar.gz

[6] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion, Technical Report No. 7*, ISRN TELCORDIA--2012-06+PR-0GARAU, 2012.